IN THE UNITED STATES DISTRICT COURT

FOR THE EASTERN DISTRICT OF VIRGINIA

Alexandria Division

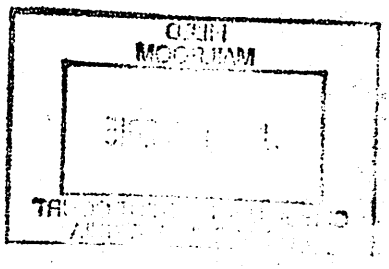
BMG RIGHTS MANAGEMENT (US) LLC, *et al.*,	*	
Plaintiffs,	*	
	*	Civil Action No. 1:14cv1611 (LO/JFA)
v.	*	
	*	
COX ENTERPRISES, INC., *et al.*,	*	
Defendants.	*	

REPLY TO THE PLAINTIFFS' OPPOSITION TO THE MOTION TO QUASH OR MODIFY

I received a letter from my ISP regarding a subpoena, which included a copy of the Order Granting Plaintiff's Application for Leave to Take Discovery.

From accounts of previous defendants of BMG Rights Management, these subpoena notifications are followed by demand letters. These letters -- which the Plaintiffs admit to demand $20 for each infraction (totaling several thousand in some cases) to avoid dealing with their lawsuit -- and their phone calls, which are persistent, are the reason I am filing this reply, and for this reason, I respectfully request that I be allowed to do so without revealing my personally identifying information.

John Doe is appearing *pro se* and is a non-party in the above-captioned case. A copy of thisreply is being provided to the attorney who issued this subpoena.



REPLY

I, John Doe, respectfully asks the court to pardon his lack of legal knowledge and lack of understanding regarding both the terminology and formatting of legal documents.

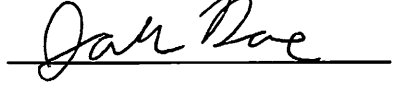
I feel that it is important for the court to know that the Plaintiffs Opposition to my Motion to Quash is deeply flawed on multiple levels.

1. The IP address 70.181.75.196 is not mine. Their information gathering is so poor, they assigned it to me improperly in a legal document submitted to this court.
2. Their information gathering is flawed and poor period. I have attached the paper *Challenges and Directions for Monitoring P2P File Sharing Networks by* Michael Piatek, Tadayoshi Kohno, Arvind Krishnamurthy. It describes the problems with monitoring Bittorrent clients and how it often leads to mis-identifying IP addresses.

WHEREFORE, John Doe respectfully prays that the Court enter an order quashing or modify said subpoena *duces tecum.*

Dated June 11, 2015

Respectfully submitted,



John Doe

Pro se

CERTIFICATE OF SERVICE

I hereby certify that on 6/11/2015, I served a copy of the foregoing document, via US Mail, on:

Jeremy D. Engle (vsb #72919)
STEPTOE & JOHNSON, LLP
1330 Connecticut Avenue, NW
Washington, DC 20036
Tel.: (202)429-3000
Fax: (202)429-3902
Attorneysfor Plaintiffs

Challenges and Directions for Monitoring P2P File Sharing Networks
– or –
Why My Printer Received a DMCA Takedown Notice

Michael Piatek* Tadayoshi Kohno * Arvind Krishnamurthy*

Abstract— We reverse engineer copyright enforcement in the popular BitTorrent file sharing network and find that a common approach for identifying infringing users is not conclusive. We describe simple techniques for implicating arbitrary network endpoints in illegal content sharing and demonstrate the effectiveness of these techniques experimentally, attracting real DMCA complaints for nonsense devices, e.g., IP printers and a wireless access point. We then step back and evaluate the challenges and possible future directions for pervasive monitoring in P2P file sharing networks.

1 Introduction

Users exchange content via peer-to-peer (P2P) file sharing networks for many reasons, ranging from the legal exchange of open source Linux distributions to the illegal exchange of copyrighted songs, movies, TV shows, software, and books. The latter activities, however, are perceived as a threat to the business models of the copyright holders [1].

To protect their content, copyright holders police P2P networks by monitoring P2P objects and sharing behavior, collecting evidence of infringement, and then issuing to an infringing user a so-called *Digital Millennium Copyright Act (DMCA) takedown notice*. These notices are formal requests to stop sharing particular data and are typically sent to the ISPs corresponding to the IP addresses of allegedly infringing users.

The combination of large-scale monitoring of P2P networks and the resulting DMCA complaints has created a tension between P2P users and enforcement agencies. Initially, P2P designs were largely managed systems that centralized key features while externalizing distribution costs, e.g., Napster's reliance on a centralized index of pointers to users with particular files. Legal challenges to these early networks were directed towards the singular organization managing the system. In contrast to these managed *systems*, currently popular P2P networks such as Gnutella and BitTorrent are decentralized *protocols* that do not depend on any single organization to manage their operation. For these networks, legal enforcement requires arbitrating disputes between copyright holders and P2P users directly.

*Dept. of Computer Science and Engineering, Univ. of Washington. E-mails: piatek@cs.washington.edu, yoshi@cs.washington.edu, arvind@cs.washington.edu. Additional information about this paper is available at http://dmca.cs.washington.edu/.

The focus of this paper is to examine the tension between P2P users and enforcement agencies and the challenges raised by an escalating arms race between them. We ground this work in an experimental analysis of the methods by which copyright holders currently monitor the BitTorrent file sharing network. Our work is based on measurements of tens of thousands of BitTorrent objects. A unique feature of our approach is that we intentionally try to receive DMCA takedown notices, and we use these notices to drive our analysis.

Our experiments uncover two principal findings:

- Copyright holders utilize inconclusive methods for identifying infringing BitTorrent users. We were able to generate hundreds of DMCA takedown notices for machines under our control at the University of Washington that were not downloading or sharing any content.
- We also find strong evidence to suggest that current monitoring agents are highly distinguishable from regular users in the BitTorrent P2P network. Our results imply that automatic and fine-grained detection of monitoring agents is feasible, suggesting further challenges for monitoring organizations in the future.

These results have numerous implications. To sample our results, based on the inconclusive nature of the current monitoring methods, we find that it is possible for a malicious user (or buggy software) to implicate (frame) seemingly any network endpoint in the sharing of copyrighted materials. We have applied these techniques to frame networked printers, a wireless (non-NAT) access point, and an innocent desktop computer, all of which have since received DMCA takedown notices but none of which actually participated in any P2P network.

Based on these observations, we then explore how the arms race between content consumers and monitoring organizations might evolve and what challenges would arise for both parties. We explicitly do not take sides in this arms race. Rather, we take special care to be independent and instead consider methods by which both users and monitoring organizations could advance their interests. Our goal is to provide a foundation for understanding and addressing this arms race from both perspectives. While couched in the context of the sharing of copyrighted content, we also believe that our results and directions will become more broadly applicable as new uses for P2P file sharing networks evolve.

Trace	Complaint type						Totals	
	Movie	Music	Television	Software	Books	Mixed	Complaints	Swarms obs.
August, 2007	82	0	11	18	11	0	122	55,523
May, 2008	200	0	17	46	0	18	281	27,545

Table 1: DMCA takedown notices received during our BitTorrent experiments. All are false positives.

2 Background

BitTorrent overview: BitTorrent is a P2P file distribution tool designed to replace large file downloads over HTTP. Rather than downloading a large file directly, a BitTorrent user instead downloads a small torrent file which contains metadata regarding the original file(s), e.g., names and sizes, as well as the address of a coordinating *tracker* for the swarm. The tracker is a rendezvous service for peers in a particular swarm, providing a random set of active downloaders upon request. New users register with the tracker, advertising their status as a potential peer, and connect to the set of peers returned by the tracker to begin exchanging data. BitTorrent peers distribute small blocks that comprise the original file. Ideally, a user with a complete copy of the file need only send each block to a few peers and the rest of the distribution will be performed by the swarm.

DMCA Enforcement: At present, DMCA takedown notices are the principle mechanism used for enforcing copyright on the Internet in the United States. DMCA notices are sent to ISPs when monitoring agencies detect alleged infringement. Separate and less frequently used mechanisms are actual legal prosecutions and "pre-settlement" letters that inform users of plans for prosecution if a settlement payment is not made. To date, we have not received any pre-settlement letters as a result of our experiments.

Takedown notices generally include the date and time of an observation, metadata for the infringing file, and the IP address of the infringing host. Network operators then respond to the complaint, often forwarding it (if possible) to the user identified by the network information.

A key question for understanding the enforcement process is: how are infringing users identified? We consider two options for detection in BitTorrent:

- *Indirect detection* of infringing users relies on the set of peers returned by the coordinating tracker only, treating this list as authoritative as to whether or not IPs are actually exchanging data within the swarm.
- *Direct detection* involves connecting to a peer reported by the tracker and then exchanging data with that peer. Direct detection has relatively high resource requirements, a topic we revisit in Section 6.

While direct detection is more conclusive and is the stated approach for monitoring the Gnutella P2P network by at least one content enforcement agency [11], we find that many enforcement agencies instead use indirect detection when monitoring BitTorrent.

3 Data Sources and Methodology

Our understanding of copyright enforcement in BitTorrent is based on measurement and analysis of tens of thousands of live BitTorrent swarms and the DMCA complaints these measurements attracted. To gather a set of candidate swarms to monitor, we continuously crawled popular websites that aggregate torrent metadata. For each observed swarm, our instrumented BitTorrent clients contacted the associated tracker, requesting a set of bootstrapping peers. These requests were repeated for each swarm every 15 minutes from 13 vantage points at the University of Washington. Crucially, querying the tracker for a set of bootstrapping peers allowed us to determine membership in swarms and advertise our presence as a potential replica *without uploading or downloading any file data whatsoever.*

The process of collecting these traces generated many DMCA takedown notices; these are summarized in Table 1. Our initial trace (August, 2007) was collected in support of a separate measurement study of BitTorrent [9]. During this prior work, we viewed DMCA complaints as an annoyance to be avoided. More recently, the realization that we had managed to attract complaints without actually downloading or uploading any data prompted us to revisit the issue. Analyzing the complaints in more detail, we were surprised to find multiple enforcement agencies sourcing takedown notices for different content, demonstrating that spurious complaints (for machines that were not actually infringing) were not isolated to a single agency (or industry).

In May, 2008, we conducted a new measurement study of BitTorrent aimed at answering two questions. First, *has the enforcement approach changed?* We find that it has not; we continue to receive DMCA complaints even in the absence of data sharing. Our second question is: *can a malicious user falsely implicate a third party in copyright infringement?* We find that framing is possible given the monitors' current use of indirect detection of infringing users, a topic we discuss next.

4 False Positives with Indirect Detection

The main weakness in current methods of detecting copyright infringement in BitTorrent appears to be the treatment of indirect reports as conclusive evidence of

Host type	Number of complaints
Desktop machine (1)	5
IP Printers (3)	9
Wireless AP (1)	4

Table 2: False positives for framed addresses.

participation. We now describe how the use of indirect reports exposes monitoring agents and innocent users to attacks from malicious users attempting to implicate others. We verify one variant of this family of attacks experimentally and quantify its effectiveness in the wild.

4.1 The Misreporting Client Attack

The first request from a BitTorrent client to a tracker serves two purposes. First, it elicits a response that provides the newly joined client with an initial set of peers with which to exchange data. Second, the request notifies the tracker that a new peer is available and can be listed in responses to future requests. By default, BitTorrent trackers record the source IP address from the request as the actual address of the peer to be delivered to others. But, some BitTorrent tracker implementations support an optional extension to the peer request message that allows requesting clients to specify a different IP address that the tracker should record in its list of peers instead. This is intended to provide support for proxy servers and peers/trackers behind the same NAT. But, when combined with the lack of verification of tracker responses by monitoring agents, this extension also allows malicious clients to frame arbitrary IPs for infringement via a simple HTTP request. We refer to this behavior as the misreporting client attack. A sample HTTP request to frame a target IP address A.B.C.D, after standard parsing of the relevant torrent metadata, is as follows:

```
wget 'http://torrentstorage.com/announce.php
?info_hash=%0E%B0c%A4B%24%28%86%9F%3B%D2%CC%
BD%0A%D1%A7%BE%83%10v&peer_id=-AZ2504-tUaIhr
rpbVcq&port=55746&uploaded=0&downloaded=0&le
ft=366039040&event=started&numwant=50&no_pee
r_id=1&compact=1&ip=A.B.C.D&key=NfBFoSCo'
```

We designed our May, 2008 experiments to examine the effectiveness of this attack in the wild today. For each tracker request issued by our instrumented clients, we included the option for manually specifying a client IP to frame, drawing this IP randomly from a pool of IPs at the University of Washington. Each framed IP was under our direct control and none were engaged in any infringing activity. These addresses include printers, a wireless access point, and an ordinary desktop machine. As a consequence of our spoofed requests, all of these devices attracted complaints (as summarized in Table 2). We also attempted to frame two IP addresses for which no machines were associated; these IP addresses were not remotely pingable and we did not receive any complaints for these IP addresses.

Although successful, the yield of misreporting client attack is low. Of the 281 complaints generated by our May, 2008 trace, just 18 of these were for IPs that we were attempting to implicate. The remaining majority were targeted at the IP addresses from which we launched our spoofed requests. Yield was low with our initial experiments because we did not know *a priori* which trackers support the protocol extension required for IP spoofing. Those that do not simply disregard that portion of the request message and instead record the IP source address of the request message. Thus, the effectiveness of the vanilla misreporting client attack, as described above, depends on what fraction of swarms can be spoofed.

We can compute this fraction using our measurements. In addition to implicating IPs continuously, we also record swarm membership continuously. Because we know that our framed IPs did not participate in BitTorrent swarms, observing *any* framed IP in the set of peers returned by a tracker indicates that the given tracker (and swarm) support spoofed addresses. Over the duration of our trace, we observed our framed IPs in 5.2% of all swarms, suggesting that the limited yield of the misreporting client attack is simply the result of a small fraction of swarms supporting spoofing as opposed to any sanity checks that might detect spoofed IPs.

More sophisticated variants of our attacks could route the HTTP requests through a proxy or anonymization service like Tor, and could also target only those trackers that support spoofed addresses.

4.2 Additional sources of false positives

Our experiments confirm that a malicious user can implicate arbitrary IPs in illegal sharing today. But, the misreporting client attack is not the only source of false positives possible given the current approach to enforcement.

Misreporting by trackers: The most straightforward way to falsely implicate an IP address in infringement is for the coordinating tracker to simply return that IP address as a peer regardless of participation. Since the torrent metadata files that specify trackers are user-generated, a malicious user can frame arbitrary IPs simply by naming his own misreporting tracker during the creation of the torrent and then uploading that torrent to one of the many public aggregation websites that we (and enforcement agencies, presumably) crawl. From the perspective of users downloading the file, such a malicious tracker would seem no different than any other.

Mistimed reports: A tracker need not be malicious to falsely implicate users. Consider the following scenario. Bob participates in an infringing BitTorrent swarm from a laptop via WiFi with an IP address assigned via DHCP, e.g., at a university or coffee shop. Bob then closes his laptop to leave, suspending his BitTorrent client with-

out an orderly notification to the tracker that he has left. Some time later, Alice joins the same WiFi network and, due to the DHCP timeout of Bob's IP, Alice receives Bob's former address. Simultaneously, a monitoring agent queries the tracker for the swarm Bob was downloading and the tracker reports Bob's former IP. The monitoring agent then dispatches a DMCA notice to the ISP running the WiFi network naming Bob's IP but with a timestamp that would attribute that IP to Alice, a false positive. Whether this is a problem in practice depends on the relative timeouts of BitTorrent trackers and DHCP leases, neither of which is fixed. In a university environment in 2007, DHCP lease times were set to 30 minutes [4]. The interarrival time of tracker requests is typically 15 minutes at least, meaning that even a conservative tracker timeout policy of two missed requests coupled with a 30 minute DHCP lease time could result in this type of misidentification.

Man-in-the-middle: Because BitTorrent tracker responses are not encrypted, man-in-the-middle attacks at the network level are straightforward. Anyone on the path between tracker and a monitoring agent can alter the tracker's response, implicating arbitrary IPs. Further, man-in-the-middle attacks are also possible at the overlay level. For redundancy, current BitTorrent clients support additional methods of gathering peers beyond tracker requests. These include peer gossip and distributed hash table (DHT) lookup [3]. Although we have not determined experimentally if these sources of peers are used by monitoring agents, each permits man-in-the-middle attacks. DHT nodes can ignore routing requests and return false IPs in fraudulent result messages. Similarly, peers can gossip arbitrary IPs to their neighbors.

Malware and open access points: There are other ways in which innocent users may be implicated for copyright infringement. For example, their computer might be running malware that downloads or hosts copyrighted content, or their home network might have an open wireless access point that someone else uses to share copyrighted content. We do not consider these further in this paper since, in these cases, the user's IP address is involved in the sharing of copyrighted content (even if the user is innocent). Our previous examples show how it is possible for a user's IP address to be incorrectly accused of copyright violation even if no computer using that IP address is sharing copyrighted content at the time of observation.

5 False Negatives with Direct Detection

A common method employed by privacy conscious users to avoid systematic monitoring is IP blacklists. These lists include the addresses of suspected monitoring agents and blacklisting software inhibits communication to and from any peers within these address ranges. The popularity of blacklists is, in retrospect, perhaps a bit surprising given our discovery (Section 4) that monitoring agents are issuing DMCA takedown notices to IP addresses without ever exchanging data with those IPs. Nevertheless, blacklists—if populated correctly—might be effective in protecting against direct monitoring techniques that involve actual data exchange between monitoring agents and P2P clients.

Since we expect that enforcement agencies will soon shift to more conclusive methods of identifying users, we revisit the issue of blacklists and ask: if enforcement depended on direct observation, are current blacklists likely to inhibit monitoring? We find that the answer to this question is likely no; current IP blacklists do not cover many suspicious BitTorrent peers. In this section, we describe the trace analysis supporting this conclusion.

In considering which peers are likely monitoring agents and which are normal BitTorrent users, our main hypothesis is that current monitoring agents are crawling the network using methods similar to our own; i.e., crawling popular aggregation sites and querying trackers for peers. On our part, this behavior results in our measurement nodes appearing as disproportionately popular peers in our trace, and systematic monitoring agents are likely to exhibit similarly disproportionate popularity.

To test this, we first define our criteria for deciding whether or not a peer is likely to be monitoring agent, beginning by considering the popularity of peers observed in our trace on a single day (May 17th, 2008). Of the 1.1 million reported peers in 2,866 observed swarms, 80% of peers occur in only one swarm each. Of the remaining 20% that occur in multiple swarms, just 0.2% (including our measurement nodes and framed IPs) occur in 10 or more swarms. The disproportionate popularity of this small minority suggests the potential for measurement agents, but manual spot-checks of several of these IPs suggests that many are ordinary peers; i.e., they come from addresses allocated to residential broadband providers and respond to BitTorrent connection requests.

Other addresses, however, come from regions allocated to ASes that do not provide residential broadband, e.g., co-location companies that serve business customers only. Further, in several instances multiple addresses from the /24 prefixes of these organizations are among the most popular IPs and none of the addresses respond to BitTorrent connection requests. We take this as a strong signal that these are likely monitoring agents and consider any /24 prefix with six or more hosts listed in ten or more swarms to be suspicious. We manually inspected the organization information for these IPs (using whois lookup), eliminating any ASes that provide residential service. Although these ASes may host monitoring agents, we adopt a conservative standard by discarding them. This further pruning resulted in a set of 17

suspicious prefixes.

To test our list of suspicious prefixes against blacklists, we obtained the latest versions of blacklists used by the popular privacy protection software SafePeer and PeerGuardian. Of the 17 suspicious prefixes, 10 were blocked, and 8 of these, while allocated to a co-location service provider, are attributed in the blacklists to either MediaSentry or MediaDefender, copyright enforcement companies. However, seven of our suspicious prefixes (accounting for dozens of monitoring hosts) are not covered by current lists.

Repeating this analysis for additional days of our trace yields similar results, suggesting that existing blacklists might not be sufficient to help privacy conscious peers escape detection (possibly because these blacklists are manually maintained). On the other hand, our analysis also implies monitoring agents could be automatically detected by continuously monitoring swarm membership and correlating results across swarms. While the exact behavior of future monitoring peers may change, we posit that their participation in swarms will remain distinguishable. Adoption of detection techniques like ours would make it harder for monitoring agencies to police P2P networks without exposing themselves, an issue we elaborate on in the next section.

6 Lessons and Challenges

The current state of P2P monitoring and enforcement is clearly not ideal. The potential for false positives and implication of arbitrary addresses undermines the credibility of monitoring and creates a significant inconvenience for misidentified users (if not financial and/or legal penalties). We now discuss the implications of our work, considering lessons learned and likely future challenges for each of the principals involved in copyright enforcement: enforcement agencies, ISPs, and users.

6.1 Enforcement agencies

The main lesson for enforcement agencies from our work is that new methods of collecting user information are required for identification to be conclusive. A more thorough approach to detecting infringement in BitTorrent would be to adopt the stated industry practice for monitoring the Gnutella network: in the case of suspected infringement, download data directly from the suspected user and verify its contents [11]. Because we have notified several enforcement agencies of the vulnerabilities described in Section 4, we expect increasing use of direct downloads for verifying participation. This reduces the potential for false positives, but it is likely to significantly increase the cost of enforcement as well as the risk of exposing monitoring agents.

The cost of direct identification: The current monitoring approach for BitTorrent, simply issuing a tracker request, requires only a single HTTP request and response, generating at most a few kilobytes of network traffic, a single connection, and minimal processing. In contrast, directly connecting to users and downloading data would require a TCP connection apiece for each potential peer, block transfers (blocks are typically hundreds of kilobytes), and hash computations to verify data integrity.

This translates into a 10-100X increase in the throughput required for monitoring swarms. Our August, 2007 crawl, which relied primarily on tracker requests, required roughly 100 KBps of sustained throughput per measurement node to monitor roughly 55,000 swarms crawled over the course of a month. For a period of one month, direct verification of our trace would require 25 terabytes of traffic as compared to just 2.5 terabytes for indirect monitoring. Furthermore, verifying participation by directly downloading data from peers is only possible for those peers that are not masked by NATs or firewalls. Detecting those that are requires sustained operation as a server; i.e., waiting for connection requests, accepting them, and then engaging in transfers to confirm participation, further increasing the complexity and resources required for large-scale, direct monitoring.

The risk of exposing monitoring agents: A major challenge for enforcement agencies is coverage; i.e., identifying all infringing users. From the perspective of monitoring agents, achieving high coverage is straightforward; simply crawl and monitor all swarms. From the perspective of coordinating trackers, however, this behavior amounts to a denial of service attack. Many swarms are hosted on a small number of public trackers. Monitoring agents that issue frequent requests for each of the thousands of swarms that one of these public trackers coordinates are likely to be detected and blocked. Indeed, our own monitors were blocked from several of these trackers prior to rate-limiting our requests.

To avoid notice today, monitoring agents need to acquire multiple IPs in diverse regions of the address space and limit their request rate. But, IP addresses are an increasingly scarce (and expensive) resource, and monitoring more than a few swarms daily from each IP risks exposing monitoring agents through their disproportionate popularity. Given these challenges, recent calls from industry to enlist ISPs directly in enforcement are unsurprising [7]. Since ISPs do not need to participate in P2P networks to monitor user behavior, there are no apparent monitoring agents to block. The majority of complaints we have received to date reflect the tradeoff between coverage and exposure; they primarily target recently released movies, DVDs, or software packages, even though we appeared to download many more old works than new.

Challenges to direct monitoring: Even if a monitoring

agent connects directly to a device behind a given IP address, there are challenges to associating the endpoint of that communication directly to a specific physical machine, let alone a specific user. For example, suppose the IP address corresponds to a family's home cable-modem or DSL connection, and suppose the family has an open wireless access point (or an insecurely-protected access point) on their internal network. It may be challenging to determine whether the machine participating in the P2P network belongs to the family or a neighbor. To address this challenge, monitoring agents may in the future collect data about not only the IP addresses of potentially infringing parties but also operating system [8, 10, 12] and physical device [5] fingerprints.

6.2 ISPs

For ISPs, the main lesson from our work is that sanity checking is necessary to protect users from spurious complaints but not sufficient. Section 4 details several scenarios which may result in false positives that can be detected by diligent network operators. However, not all false positives can be detected, and current trends in enforcement are towards increased automation rather than increased sanity checking of complaints.

Increasing automation: Because most DMCA complaints are communicated over email, network operators typically inspect messages manually to identify users. At the University of Washington, this manual step has served as an important check that eliminates some erroneous complaints before they reach users [2].

Although having a human "in the loop" is beneficial to users, it may not be tenable with increasing rates of enforcement. While we continuously monitored tens of thousands of swarms in our traces, we garnered only hundreds of complaints, a small fraction of potentially infringing swarms. Even at this limited level of enforcement, many universities still require dedicated staff to manually process all the complaints sent to their users, increasing costs. Enforcement agencies rely on cooperation from network operators to identify infringing users, but increasing costs have pushed both ISPs and monitoring agencies towards automated enforcement.

The trend towards automation is reflected in the properties of complaints themselves. The delay between the observation of peers by enforcement agencies and the timestamp of complaint email messages has reduced significantly. The median delay for complaints generated by our trace from August, 2007 is 49 hours. For more recent complaints collected in May, 2008, the median delay is just 21 hours. Further, these recent complaints increasingly include machine-readable summaries of their content, e.g., XML data with public schemas. We hypothesize that the intent is to automate the complaint process at the levels of both enforcement agency and ISP. Enforcement agencies can crawl P2P networks, generating and dispatching XML complaints which can then be parsed by ISPs and automatically forwarded to users with no human intervention.

6.3 Users

Our results show that potentially any Internet user is at risk for receiving DMCA takedown notices today. Whether a false positive sent to a user that has never even used BitTorrent or a truly infringing user that relies on incomplete IP blacklists, there is currently no way for anyone to wholly avoid the risk of complaints. But, the current approach to enforcement has a natural limiting factor. To avoid being detected, our traces suggest that enforcement agents are not monitoring most swarms and tend to target those new, popular swarms that are the most economically valuable.

In the long term, the main challenge for privacy conscious users is to develop a way to systematically detect monitoring agents. We consider two cases. If enforcement agencies continue to monitor swarms at the *protocol level* by participating in swarms, users may develop new techniques to build more dynamic, comprehensive blacklists. If ISPs are enlisted in enforcement at the *network level* by collecting traces of user traffic, we anticipate increased use of stronger encryption to frustrate realtime, automated identification of P2P protocols. We expand on each of these in turn.

Blacklists on-the-fly: Just as we expect enforcement agencies to shift from indirect to direct methods of enforcement, we also expect P2P developers to evolve IP blacklisting techniques. Currently, blacklists are centrally maintained and updated without systematic feedback from P2P users, ignoring a rich source of data: the observations of users. Many P2P networks include explicit mechanisms to identify and reward "good users"; e.g., tit-for-tat mechanisms reward contributions in BitTorrent and eDonkey. Future P2P networks may employ similar mechanisms to identify monitoring agents, gossiping this information among peers. Our traces show that the properties of monitoring agents today make this a straightforward task: they appear to share no data whatsoever, occur frequently in swarms, and are drawn from a small number of prefixes. Alternatively, sophisticated users may also try to generate honeypots (much like our own) that do not infringe or aid in copyright infringement, but that will be better able to detect (and hence dissuade) spurious DMCA takedown notices and coordinated monitoring.

Stronger encryption: Today, some BitTorrent clients include an option to use weak encryption to frustrate the traffic shaping methods used by several ISPs [6]. In the future, this encryption might be strengthened. For example, a tracker might assist two peers in establishing a

shared key in the face of ISPs that would otherwise attempt to identify and restrict P2P traffic. Such a tracker could include not only the IP addresses of participating clients, but also one-time public keys to decrease exposure to inline man-in-the-middle cryptographic attacks. To further resist monitoring, communications with trackers would have to be authenticated as well, perhaps by leveraging a lightweight, distributed PKI with popular trackers as the root authorities.

7 Conclusion

Although content providers are increasingly relying on systematic monitoring of P2P networks as a basis for deterring copyright infringement, some currently used methods of identifying infringing users are not conclusive. Through extensive measurement of tens of thousands of BitTorrent swarms and analysis of hundreds of DMCA complaints, we have shown that a malicious user can implicate arbitrary network endpoints in copyright infringement, and additional false positives may arise due to buggy software or timing effects. We have further demonstrated that IP blacklists, a standard method for avoiding systematic monitoring, are often ineffective given current identification techniques and provide only limited coverage of likely monitoring agents. These observations call for increased transparency and openness in the monitoring and enforcement process and build our understanding of current challenges and potential next steps for all parties involved in P2P file sharing: enforcement agencies, ISPs, and users.

8 Acknowledgments

We thank Ed Lazowska, Erik Lundberg, Scott Rose, Daniel Schwalbe, and Voradesh Yenbut. This work is supported by the NSF grants CNS-0720589, CNS-0722000, CNS-0722004 and by the University of Washington Department of Computer Science and Engineering.

References

[1] R. Cotton and M. L. Tobey. Comments of NBC Universal, Inc. In the Matter of Broadband Industry Practices. FCC Filing. WC Docket No. 07-52. http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519528962.

[2] Daniel Schwalbe. Personal communication, 2008.

[3] J. Falkner, M. Piatek, J. P. John, A. Krishnamurthy, and T. Anderson. Profiling a million user DHT. In *IMC*, 2007.

[4] M. Khadilkar, N. Feamster, R. Clark, and M. Sanders. Usage-based DHCP lease time optimization. In *IMC*, 2007.

[5] T. Kohno, A. Broido, and K. Claffy. Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, 2(2):93–108, April 2005.

[6] Message stream encryption. http://www.azureuswiki.com/index.php/Message_Stream_Encryption.

[7] MPAA wants ISP help in online piracy fight. http://news.cnet.com/8301-10784_3-9780401-7.html.

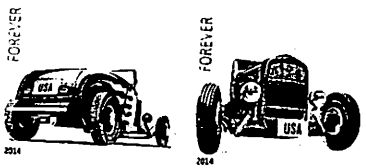
[8] Nmap - free security scanner for network exploration & security audits. http://nmap.org/.

[9] M. Piatek, T. Isdal, A. Krishnamurthy, and T. Anderson. One hop reputations for peer to peer file sharing workloads. In *NSDI*, 2008.

[10] Project details of p0f. http://freshmeat.net/projects/p0f/.

[11] C. Rampell. How it does it: The RIAA explains how it catches alleged music pirates. http://chronicle.com/free/2008/05/2821n.htm.

[12] Xprobe2. http://xprobe.sourceforge.net/.



The Honorable John F. Anderson
Albert V. Bryan US Courthouse
401 Courthouse Square
Alexandria, VA 22314